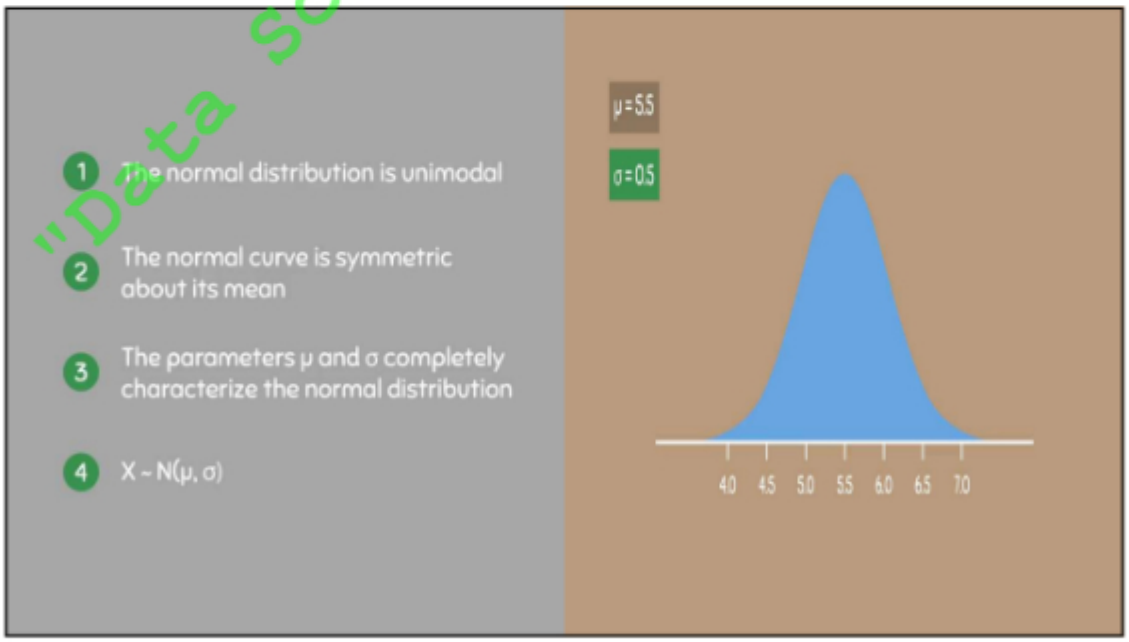
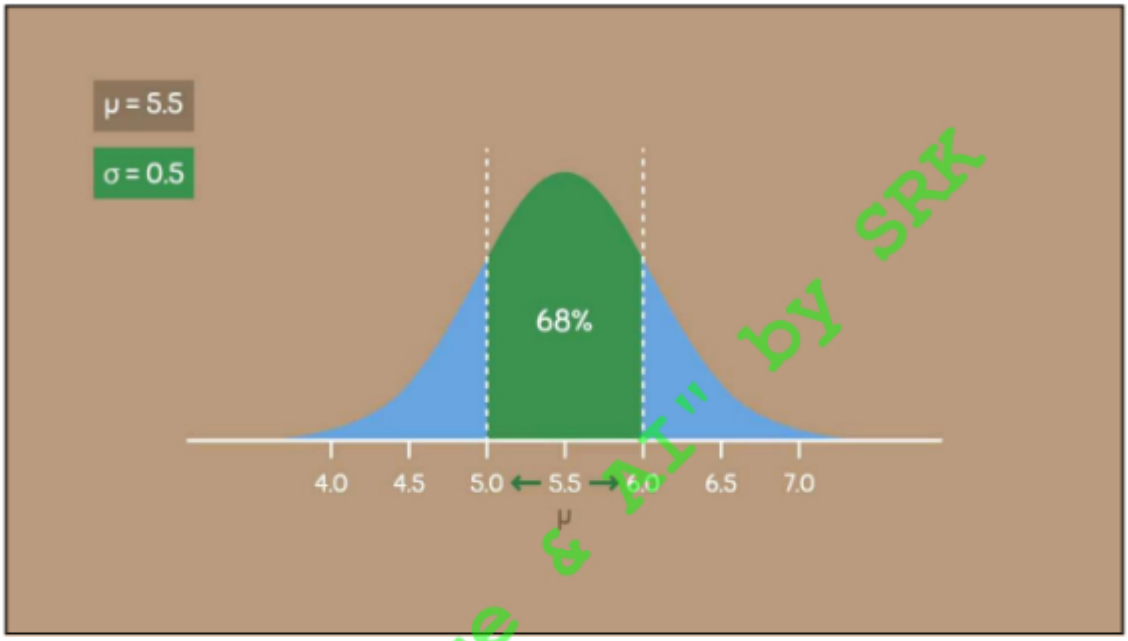
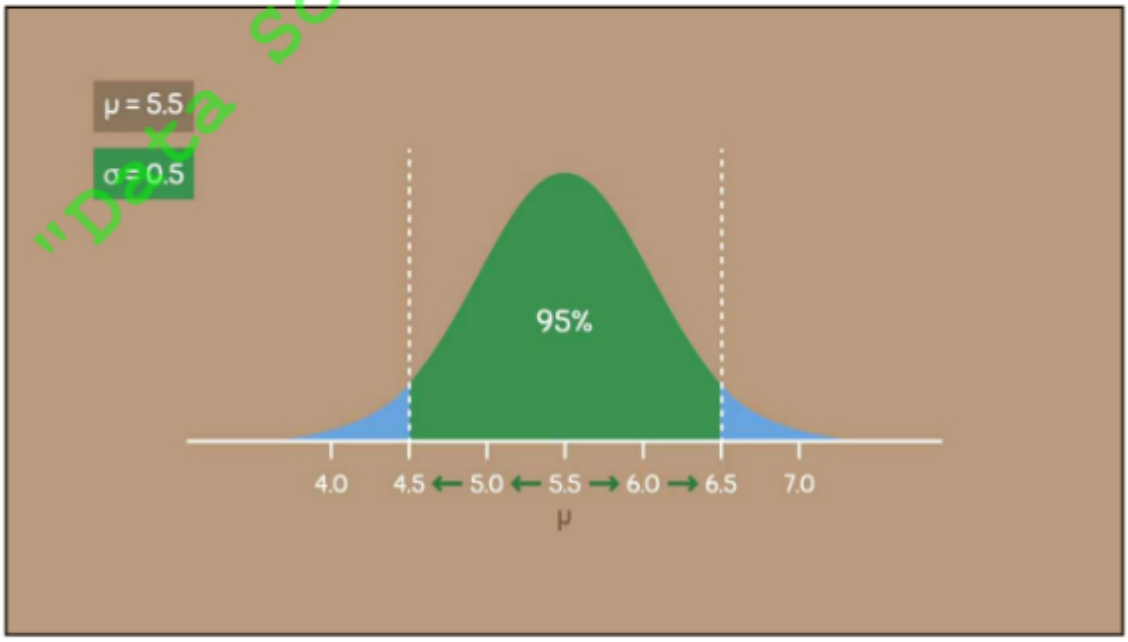
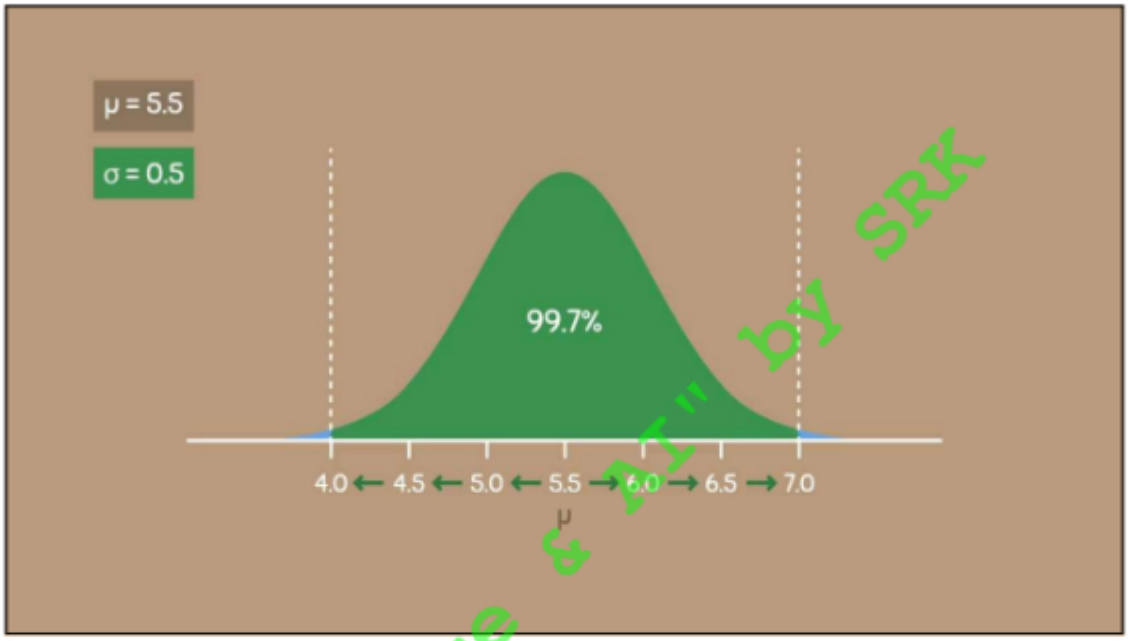
"Data Science & AI" by SRK







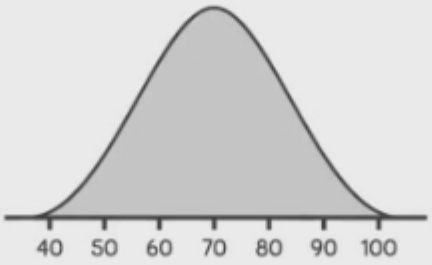
"Data Science & AI" by SRK



"Data Science & AI" by SRK

PRACTICE QUESTIONS

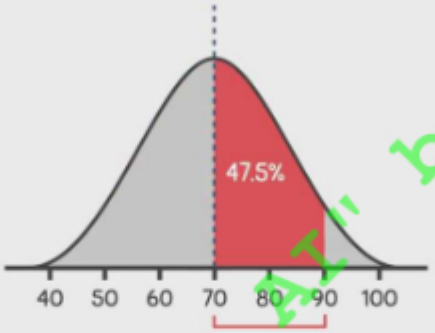
❶ The normal distribution below has a standard deviation of 10. Approximately what area is contained between 70 and 90?



PRACTICE QUESTIONS

❶ The normal distribution below has a standard deviation of 10. Approximately what area is contained between 70 and 90?

$\mu = 70$
$\sigma = 10$



"Data Science & AI" by SRK

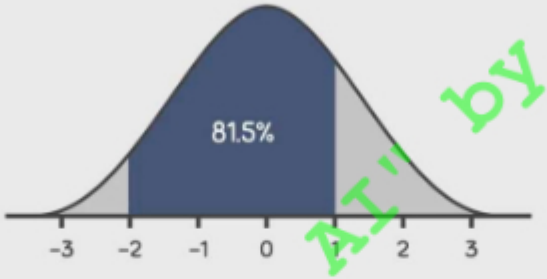
PRACTICE QUESTIONS

❷ For the normal distribution below, approximately what area is contained between −2 and 1?

PRACTICE QUESTIONS

❷ For the normal distribution below, approximately what area is contained between −2 and 1?

$\mu = 0$
$\sigma = 1$



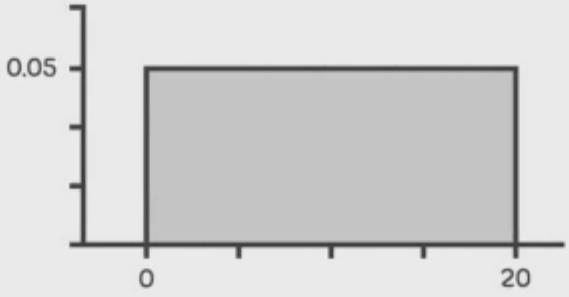
"Data Science & AI" by SRK

UNIFORM DISTRIBUTION

"Data Science & AI" by SRK

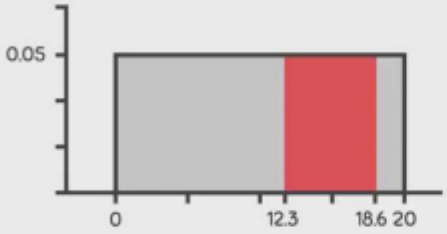
PRACTICE QUESTIONS

❸ For the uniform distribution below, what proportion of values are located between 12.3 and 18.6?



PRACTICE QUESTIONS

❸ For the uniform distribution below, what proportion of values are located between 12.3 and 18.6?



Area = L x W

= 6.3 x 0.05

Area = 0.315

"Data Science & AI" by SRK

Probability Distribution

Probability Distribution

Graphical representation of variable & respective probabilities of variable.

Types of Probability Distribution

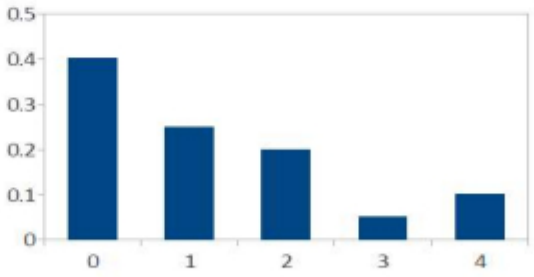
- Discrete Probability Distribution
- Continuous Probability Distribution

"Data Science & AI" by SRK

Discrete Probability Distribution

The daily sales of large flat panel TVs at a store (X)

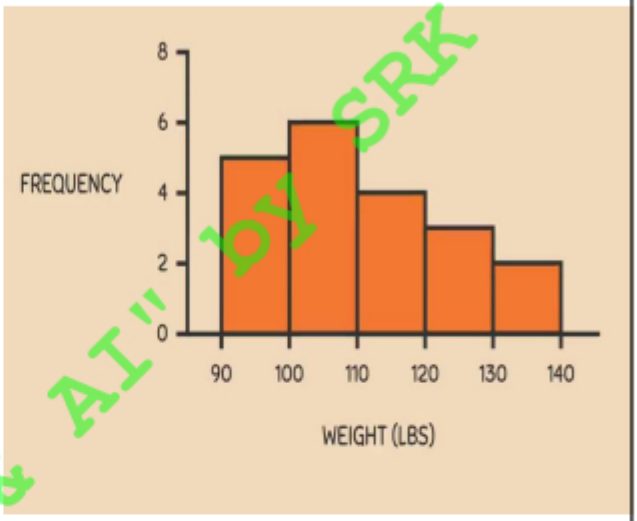
x	P(X=x)
0	0.40
1	0.25
2	0.20
3	0.05
4	0.10



What is the probability of a sale?

What is the probability of selling at least three TVs?

Continous Probability Distribution

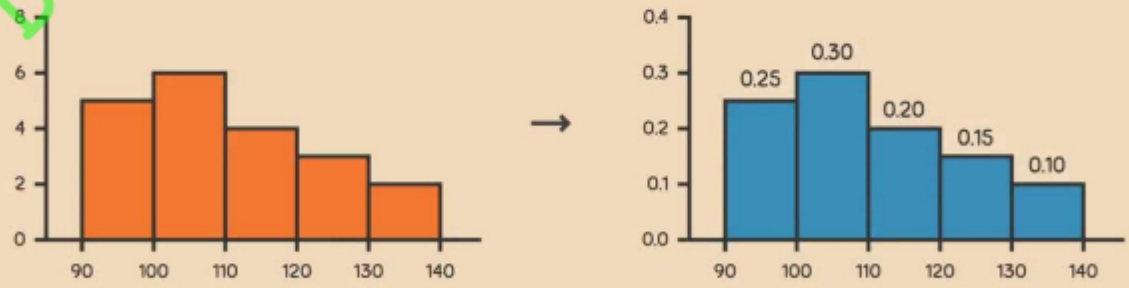


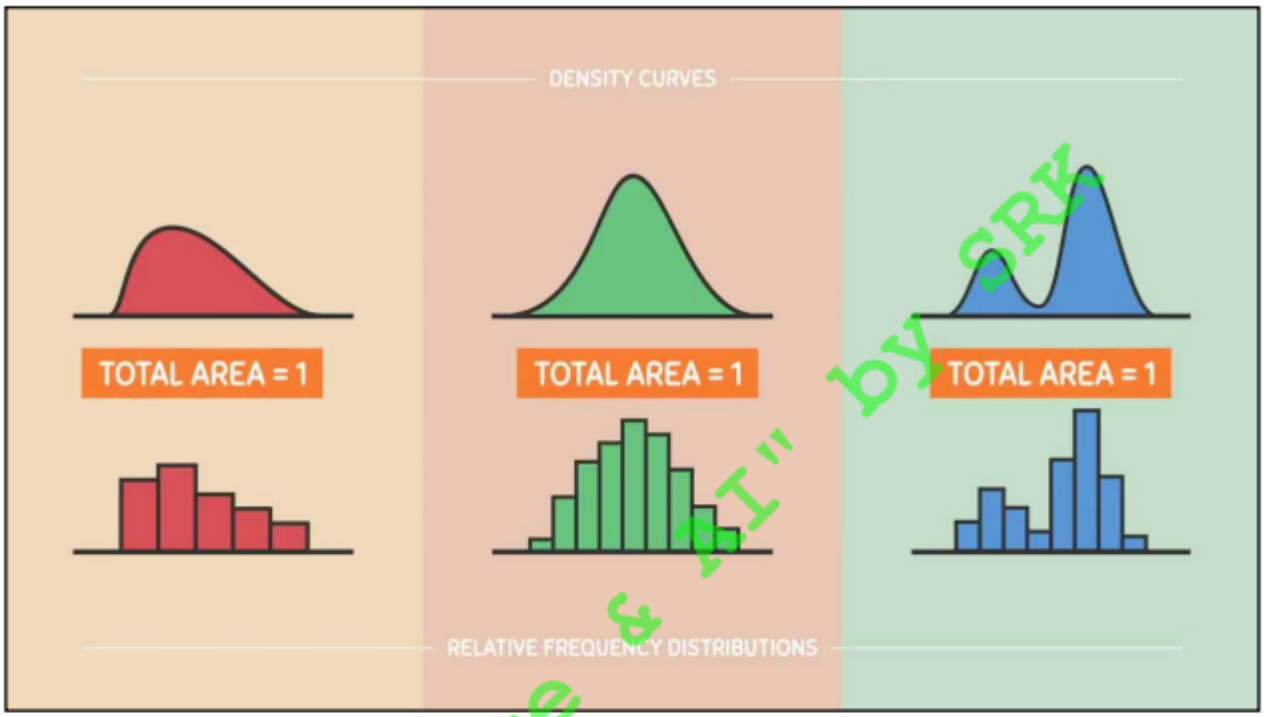
"Data Science & AI" by SRK

REVIEW

n = 20

"REGULAR" FREQUENCY DISTRIBUTION





"Data Science & AI" by SRK

Probability

Variable :

- Chance of occurrence .
- Ex: rolling a die, tossing a coin

Random Variable:

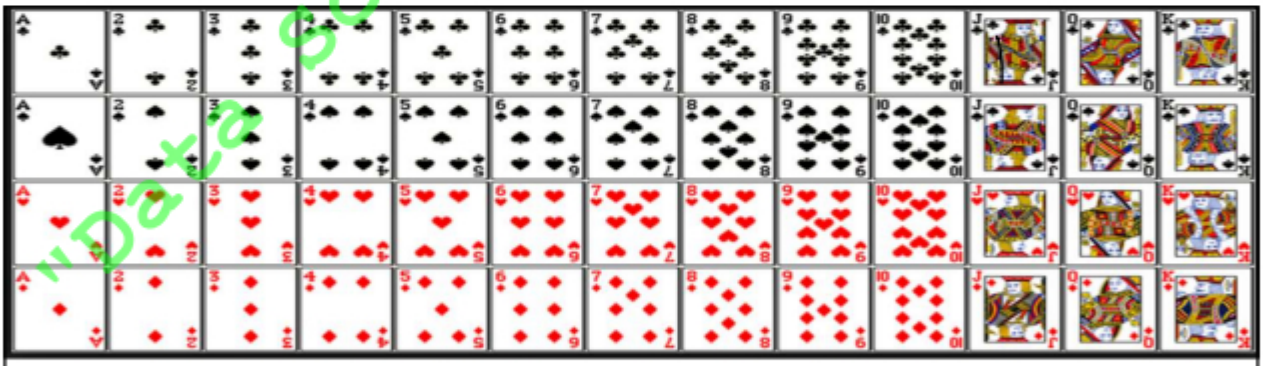
A random variable is probability associated each possibility of variable .

It is a random because there is some chance associated with each possible value.

Probability = No. of interested events/total no. of outcomes

- Always probability value lies between 0 to1.
- Sum of all Probabilities =1

"Data Science & AI" by SRK



Suppose you have randomly picked a card from the card deck. What is the probability that this card will be?

- Bigger than 10?
- Equal to or Bigger than 10?
- Smaller than 3
- Greater than 4 and less than 8

If A & B are two independent events

P(A & B) = P(A) * P(B)

Ex: probability of getting Red & 9

P(A or B) = P(A) + P(B) – P(A & B)

Ex: probability of getting Red or 9

"Data Science & AI" by SRK

COVARIANCE

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

"Data Science & AI" by SRK

CORRELATON

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X
- σ_Y is the standard deviation of Y

CORRELATON

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

"Data Science & AI" by SRK

Range	Strength of association
0	No association
0 to ±0.25	Negligible association
±0.25 to ±0.50	Weak association
±0.50 to ±0.75	Moderate association
±0.75 to ±1	Very strong association
±1	Perfect association

OUTLIER

A DATA VALUE THAT IS NUMERICALLY DISTANT FROM A DATA SET

"Data Science & AI" by SRK

OUTLIER

A DATA VALUE THAT IS NUMERICALLY DISTANT FROM A DATA SET

THE MEAN IS AFFECTED BY THE PRESENCE OF OUTLIERS

A DATA VALUE IS CONSIDERED TO BE AN OUTLIER IF..

DATA VALUE < Q1 – 1.5(IQR)

OR

DATA VALUE > Q3 + 1.5(IQR)

"Data Science & AI" by SRK

FIVE NUMBER SUMMARY

10 25 33 36 59

10 11 12 25 25 27 31 33 34 34 35 36 43 50 59

A DATA VALUE IS AN OUTLIER IF IT IS

LESS THAN 8.5

GREATER THAN 52.5

IQR = 11

"Data Science & AI" by SRK

FIVE NUMBER SUMMARY

10 25 33 36 59

10 11 12 25 25 27 31 33 34 34 35 36 43 50 59